DOCUMENT RESUME

ED 169 774 FL 010 247

AUTHOR Bialystok, Ellen; Howard, Joan
TITLE Inferencing as an Aspect of Cloze Test Performance. Working Papers on Bilingualism, No. 17.
INSTITUTION Ontario Inst. for Studies in Education, Toronto. Bilingual Education Project.
SPONS AGENCY Department of the Secretary of State, Ottawa (Ontario).
PUB DATE Apr 79
NOTE 12p.
AVAILABLE FROM Bilingual Education Project, The Ontario Institute for Studies in Education, 252 Bloor St. West, Toronto, Ontario, Canada M5S 1V6 (as long as supply lasts.)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Advance Organizers; *Cloze Procedure; *Cognitive Processes; Context Clues; French; *Language Instruction; *Language Processing; *Language Proficiency; *Language Research; Language Skills; *Language Tests; Logical Thinking; Reading Comprehension; Reading Development; Reading Skills; Secondary Education; Second Language Learning; Skill Development; Teaching Methods
IDENTIFIERS *Inference

ABSTRACT

This study investigates the role of inferencing in cloze test performance and the factors that facilitate that inferencing. Four groups of high school students learning French as a second language completed sets of cloze passages under four treatment conditions. Three conditions provided a potential cue to inferencing, while the fourth was a control condition with no additional aids. The results indicated significant differences in performance attributable to the four classes, the position of the story in the set of passages, and the inferencing treatment condition. A more detailed analysis examined the relationship between experimental condition, type of error committed, and the form class of the original word required by the text. Differences in these patterns were apparent, although non-significant. It was found that certain inferencing cues can have a facilitating effect on students' cloze performance although the pattern of facilitation is different from that found on other criterion tasks. Furthermore, it is possible to encourage students' inferencing behavior through classroom training, thereby improving cloze performance. (Author/AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ellen Bialystok
Research Officer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED169774

Inferencing as an Aspect of Cloze Test Performance

Ellen Bialystok and Joan Howard

The Ontario Institute for Studies in Education

Attention in the area of second language testing has recently been focused on the use of cloze test procedures as a measure of general L2 proficiency. The current popularity of this procedure is largely due to two factors: first, the ease of administration and subsequent scoring for large groups of subjects because of its paper-and-pencil format; and second, the nature of the cloze test as an integrative measure of ability rather than a discrete-point testing measure (Oller, 1973b).

In a cloze test, subjects are presented with a prose passage in the target language in which every nth word, usually every 7th or 8th word, has been deleted and are then asked to complete the passage by inserting the appropriate word in each blank. This task is considered to incorporate aspects of both receptive and productive abilities since the respondent must actively participate in the text construction in order to fill in the correct missing item. Both the linguistic redundancies and the contextual information provided by the passage produce for the respondent what Oller (1973a) has called a 'grammar of expectancy'. This gauge of expectancy or probability allows him to make certain assumptions about the nature of the omitted language element which are then either confirmed or rejected after continued reading of the prose passage.

The validity of the cloze method has already been demonstrated by several studies which show high correlations between performance on cloze tests and other standardized measures of L2 proficiency. Oller & Conrad (1971), working with adult learners of English as a second language, report a high correlation between students' performance on a cloze test and results on the reading comprehension (.90) and dictation (.82) subtests of the UCLA English as a Second Language Placement Examination (ESLPE). The authors conclude that the cloze method is a very promising device for measuring ESL proficiency. Again with students of English as a second language, Enkvist (1977) found correlations between cloze and specialized tests of L2 ability: reading comprehension (.70 -.85), standardized intelligence tests (.70), and predictive tests on reading and verbal skills (.50 -.85).

In the area of French as a second language, Swain, Lapkin & Barik (1976) in a pilot study report correlations of .77 between French cloze scores and French achievement scores with young children at the grade 4 level. The same study reports correlations ranging from .58 -.70 between English language achievement and English cloze data. A follow-up study (Lapkin & Swain, 1977) again reports correlations of .52 -.68 between

FL010247

English cloze scores and English language proficiency scores.

Given such consistently high correlations between cloze tests and other general language ability measures, as well as the great practical advantages offered by the cloze procedure, its potential value for classroom testing is undeniable.

However, in spite of the ample demonstration of cloze test reliability as given by the correlations with numerous other proficiency measures, the precise skills measured by the cloze test and the problem-solving processes which they presuppose have not been specified. Oller and Conrad (1971) acknowledge this deficiency but also pose the question: "Is it necessary to know exactly what a test is a test of in order to make use of it?" (p. 187). They proceed to respond to that question in the negative.

It is possible, however, that part of the process involved in solving cloze tests and consequently the skills required for their solution, can be identified. One language skill which has been hypothesized to be relevant to the solution of cloze tests is the strategy of inferencing, that is, the ability to maximally exploit available information sources to arrive at new insights into unknown aspects of the target language (Bialystok, 1978). If inferencing is indeed involved in solving cloze tests, then factors that facilitate inferencing should also facilitate cloze test performance. This hypothesis was tested in the present study.

In an earlier study, the relative efficacy of various cues as facilitators for inferencing ability was tested through students' performance on a reading comprehension and a vocabulary translation task (Bialystok, manuscript). It was found that high school students learning French as a second language could take advantage of these cues to improve performance on the two tasks. Reading comprehension, as measured by the ability to answer general content questions about the passage in English, was facilitated to the greatest degree by the presence of a relevant picture and to a lesser degree by a lesson on 'how to inference' and by access to a dictionary. Lexical translation of difficult items was, however, facilitated only in the dictionary condition.

The question for the present study was to determine the extent to which these cues which facilitated the use of the inferencing strategy on those criterion tasks would also affect performance on a cloze test, that is, a complex task for which the skills required for the solution were unclear. Their positive influence on cloze solution would support the hypothesis that inferencing is an aspect of the process of solving cloze tests and that cloze performance can be modified through training in this component skill.

Method

Subjects

The subjects for the study were 108 high school students learning French as a second language. The students were drawn from four Grade 10 classes in three different schools. All schools were in predominantly English-speaking, middle-class areas of Metropolitan Toronto.

Instrument

French prose passages which had been previously designed for use in the earlier Reading Comprehension experiment were modified for use as cloze tests in the present study. The stories were adapted to a level of difficulty appropriate to Grade 10 students after consultation with classroom teachers of French.

Each story passage was approximately 150 words long. As an introduction, there were 22 words of unbroken text, followed by the deletion of every 8th word to produce a total of 15 blanks per story. The 12 short story passages were combined in groups of three to make up four test booklets.

Design and Procedures

There were four inferencing conditions in the experiment. All students in the study performed in all experimental conditions and received all conditions in the same order, that is, control or no cue, picture cue, dictionary and lesson. The lesson was necessarily presented last so that students could not apply the techniques learned in the lesson to the other experimental conditions.

In the control condition, students were not given any additional materials or instructions to assist with their inferences.

In the second condition, a picture cue summarizing the gist of the passage accompanied each story. The picture provided the student with general, non-verbal information about the overall meaning of the story.

In the third condition subjects were given specially prepared French-English dictionaries to consult for the meaning of unfamiliar words in the passage. The dictionaries, which were 50 pages long, contained about 1000 words, including all the potentially difficult words in the set of passages as well as a number of distractors.

In the fourth condition subjects were given a brief inferencing lesson by the experimenter. To show students how to use both specific contextual clues and general knowledge of the world, several cloze examples were given in English, such as: "This morning, the _____

on Highway 401 was heavy." In another example, a list of German words was presented and students were asked to guess their meaning by comparing them to familiar English words: Butter (butter), Brot (bread), Butterbrot (sandwich), Postkarte (postcard), Kamerad (friend), qualifizieren (to qualify). If students had some prior knowledge of German, Swedish cognates were substituted: sjuk (sick), hus (house), sjukhuset (hospital). Thus they were encouraged to use their general knowledge, their knowledge of the structure of their own and the target language, and the cues provided by the context of the passage to make reasonable guesses or inferences about unknown items in the passage.

Every group of subjects completed all four test booklets, but completed each booklet under a different inferencing condition. Thus, although the order of presentation of the experimental conditions was constant, different passages were read for each condition. The experimenters made two separate visits to each class; two experimental conditions were presented at each visit.

Scoring

The responses for all subjects were scored according to two methods, an "exact word" scoring method and an "acceptable" scoring procedure. In the exact word scoring method, only the one word which was identical to that in the original passage was counted as correct. In the "acceptable" scoring method, a word synonymous with the original word or which made real sense in the context of the sentence was counted as correct. To illustrate, consider the example, "le vieux _____". Although the exact replacement word was "paysan", the word "monsieur" was scored as correct in the acceptable scoring method as it is both grammatically acceptable and meaningful within the context of the passage. Acceptability decisions were made by the experimenter. Spelling errors were disregarded in both scoring methods.

The total score possible for each story was 15, and for each set, consisting of three stories, 45. Missing data for blanks within stories were scored as 0.

Results

Although scores obtained in the acceptable scoring procedure were consistently higher than those in the exact, the patterns of scores were nonetheless identical. The correlations between the two methods of scoring for each condition were as follows: control, $r = .97$ ($N = 106$); picture cue, $r = .97$ ($N = 88$); dictionary, $r = .98$ ($N = 103$);

Table 1

Mean scores for each story in each inferencing conditions

	Condition				
	Control	Dictionary	Lesson	Picture	
Story Position					Mean
1	5.67	5.09	6.06	6.93	5.94
2	5.51	5.87	5.99	7.01	6.10
3	4.30	3.51	5.80	6.50	5.03
Mean	5.16	4.82	5.95	6.81	

lesson, r = .97 (N = 90). Accordingly, only exact scores will be reported in the data analyses.

A three-way analysis of variance which examined the effects of the four classes, the position of each story within the set and the four treatment conditions was performed on the exact scores. The differences between the four classes were significant ($F(3,82) = 9.34$, $p < .001$). While three of the classes performed to the same level of proficiency (mean scores 6.95, 6.03, 6.31) the deviant class performed significantly more poorly (mean score 3.97).

Table 1 presents the scores for the three story positions and the four inferencing conditions. The main effect of story position was significant ($F(2,164) = 32.67$), the effect being caused by the lower scores produced for the third story regardless of which story appeared in that position (Newman-Keuls $p < .01$). Since the stories themselves were standardized for level of difficulty and showed no such effect in our previous research, we attribute this difference to the diminished time available for the third story.

Insert Table 1 about here

The effect of inferencing condition was also significant ($F(3,246) = 66.49$, $p < .001$). All scores were significantly different from each other: picture cue was better than lesson ($p < .01$); lesson was better than control ($p < .01$); and control was better than dictionary ($p < .05$).

There was also an interaction between condition and story position ($F(6,492) = 7.04$, $p < .001$); the various cue conditions were not equally facilitative for all the stories. The primary effect was that the performance in the conditions offering the least effective cues, that is, control and dictionary, was particularly poor for the third story. In those conditions offering the more effective cues, the picture cue and inferencing conditions, the time constraints on the third story were less problematic and overall performance was thereby improved.

A more detailed analysis was carried out on part of the data to determine if a relationship exists between the experimental condition, the type of error committed and the part of speech required by the text. The responses for the 59 students who had clearly completed test booklets under all four experimental conditions were examined for this purpose.

For each condition, responses were first classified according to the grammatical form class of the original word, that is Noun, Verb, Modifier, Pronoun, Preposition or Functor (including articles, conjunctions and negative indicators).

Responses in each form class category were then allocated to one of seven categories: (1) omission, indicates that no response was attempted; (2) exact replacement, refers to words which are identical to the original; (3) grammatical error, refers to the exact replacement word but with a grammar and/or spelling error (e.g., sait/sais); (4) synonym, refers to words of the same grammatical category and with the same or nearly the same meaning as the original; (5) form class, refers to words which belong to the same grammatical category as the original but do not fit the meaning of the passage; (6) context, refers to those words which do not belong to the same grammatical category as the original but do make sense in the context of the passage; (7) nonsense, refers to those words which neither belong to the same grammatical category as the original nor fit into the meaning of the passage.

Mean scores were calculated for each of the six parts of speech by error type for each of the four inferencing conditions. A rank ordering of these scores, however, showed no difference attributable to form class across across cue conditions. Consequently, this factor was collapsed so that the data were examined only in terms of category and cue condition. These data are presented in Table 2.

Insert Table 2 about here

Although it is not possible to perform formal statistical analyses on these data, certain patterns of scores are nonetheless evident. The differences between the proportion scores across conditions are very slight, but because of the large number of observations contributing to each proportion, these differences merit consideration. It appears to be the case, however, that the differences between the conditions demonstrated by this subset of scores based on 59 subjects is considerably less than that obtained for the whole sample.

The three categories which do show interesting trends are omission, exact replacement and nonsense. The pattern of omissions reveals the willingness of the students in the lesson and picture conditions to generate some response as compared to the more cautious approach used in the other conditions where a greater number of omissions was recorded. These guesses are often translated into correct answers, as demonstrated by the high scores for exact replacements in the picture and lesson conditions. Reciprocally, however, the highest incidence of nonsense responses are also recorded by the lesson condition, indicating that the hunches are not always very informed.

Table 2

Raw scores and proportion of total items in each category for each experimental condition

	Picture		Dictionary		Lesson		Control	
	Score	Prop.	Score	Prop.	Score	Prop.	Score	Prop.
Omission	622	.23	739	.28	566	.21	827	.31
Exact	898	.34	791	.30	838	.32	825	.31
Grammatical Error	105	.04	173	.06	143	.05	130	.05
Synonym	172	.07	135	.05	174	.07	150	.06
Form-Class	381	.15	374	.15	427	.16	334	.13
Context	11	0	11	0	9	0	6	0
Nonsense	466	.17	432	.16	511	.19	383	.14
Total		1.00		1.00		1.00		1.00

No. of blanks/condition = 45
No. of students = 59
No. of occurrences = 2655

Discussion

The effect on performance of the three intervention conditions compared to the control was that while two of them, the picture cue and lesson, improved performance, one of them, the dictionary, impeded it. The facilitation demonstrated in the picture condition is attributed to the students' improved understanding of the gist or meaning of the passage; in the lesson condition it is attributed to the technique of arriving at reasonable hypotheses about the meaning of unknown material.

The poor results in the dictionary condition are understandable since the French-English dictionary provided could not help students to fill in an unknown item in a blank space. The dictionary could only provide assistance in the case of unknown words in the passage itself, and this knowledge of specific word meaning did not help in determining what to insert in place of the deleted word. In fact, the dictionary acted more as a hindrance than a help to the students and resulted in the lowest mean score being recorded for the dictionary condition. The additional time spent in searching for lexical items in the dictionary meant that many students did not complete the entire cloze passage in the time allowed. A greater reliance on intuitive understanding of the passage than on the need for precise translation may have provided a more successful approach to the task. Furthermore, false information occasionally resulted from the dictionary search as was the case with one respondent who, misreading the word 'souffre' (from 'souffrir': to suffer) incorrectly looked up the verb 'souffler': to blow, thus noting a completely incorrect meaning for the unknown item.

The pattern of facilitation obtained for the four conditions in the present study is different from that found earlier for the comprehension of reading passages and the translation of difficult lexical items. This difference suggests a difference in processing used in the solution to these various tasks. Each of the three, that is, reading comprehension, lexical translation, and cloze, were differentially affected by the same cue conditions. However, the examination of each of these three criterion tasks in terms of the cue conditions was motivated by the same underlying concern, that is, to determine the role of inferencing in the solution to these language tasks. The different patterns of results obtained indicate that while inferencing is involved in each task, its role in the problem-solving process is different. Across all tasks, the picture cue and inferencing lesson proved to be useful means of improving comprehension. The student was equipped either with information which he could incorporate into his comprehension strategies (picture cue) or with the means for producing better strategies (lesson) and in both cases, criterion performance was improved. The dictionary

was not legitimately an inferencing device since it provided the correct answer with no active participation required by the student. This aid had different effects in the different tasks: it was obviously useful for translating lexical items; however, it was a deterrent to solving cloze passages.

It was expected that the detailed analysis would reveal specific relationships between the cue condition and the nature of facilitation that resulted. For example, it was expected that the picture cue which represented the overall meaning of the story would be more useful for the content words, such as nouns, than for the functors. Increased general comprehension of the story was not however expected to assist with the replacement of the syntactic and grammatical aspects of the passage. These patterns did not occur in the data, but since the study was not explicitly designed to test them, the results are tentative. Further research should examine these questions by systematically deleting words which conform to certain syntactic and semantic criteria and would be related in the design to particular kinds of clues designed to improve syntactic or semantic aspects of the language.

Two main conclusions follow from the study. First, inferencing is an integral component in the solution of cloze tests. While neither an adequate description nor a complete understanding of the process by which cloze tests are solved, or indeed, any integrative language task is performed, are currently available, the identification of a component of that process is an important beginning to that understanding.

Second, the differential effect of the cue conditions demonstrated again that inferencing behaviour by students of a second language is in fact modifiable by the intervention of appropriate materials and instructions. Having thus recognized the importance of the inferencing strategy in the language learner's repertoire, these considerations point to clear implications for both classroom instruction and materials development.

References

Bialystok, E. A theoretical model of second language learning. Language Learning, 1978, 28, 69-83.

Bialystok, E. Exploiting the information: Inferencing in second language reading comprehension. Manuscript.

Enkvist, N.E. & Kahonen, V. Cloze testing, some theoretical and practical aspects. In A. Zettersten (Ed.), Papers on English Language Testing in Scandinavia (Anglica Et Americana 1). Copenhagen: University of Copenhagen, 1977, 12-42.

Lapkin, S. & Swain, M. The use of English and French cloze tests in a Bilingual Program Evaluation: Validity and error analysis. Language Learning, 1977, 27, 279-314.

Oller, J.W. Jr. & Conrad, C. The cloze technique and ESL proficiency. Language Learning, 1971, 21, 183-196.

Oller, J.W. Jr. Cloze test of second language proficiency and what they measure. Language Learning, 1973a, 23, 105-118.

Oller, J.W. Jr. Discrete-point tests versus tests of integrative skills. In J.W. Oller, Jr. & J. Richards (Eds.), Focus on the learner: Pragmatic perspectives for the language teacher. Rowley, Mass.: Newbury House, 1973b, 184-199.

Swain, M., Lapkin, S. & Barik, H.C. The cloze test as a measure of second language proficiency for young children. Working Papers on Bilingualism, 1976, 11, 32-42.